High genetic diversity and weak population structuring in the Eurasian Woodcock in Hungary during spring

Gergely Schally*, Krisztián Frank, Botond Helai, Péter Fehér, Ádám Farkas, László Szemethy & Viktor Stéger

Schally, G., Farkas, Á., Szemethy, L., Institute for Wildlife Conservation, Faculty of Agricultural and Environmental Sciences, Szent István University, Páter Károly Str. 1., 2100 Gödöllő, Hungary. * Corresponding author’s e-mail: Schally.Gergely@mkk.szie.hu
Frank, K., Helai, B., Fehér, P., Stéger, V. Agricultural Biotechnology Institute, National Agricultural Research and Innovation Centre, Szent-Györgyi Albert Str. 4., 2100 Gödöllő, Hungary

Received 23 August 2017, accepted 28 March 2018

In this study, we examined the variation of eight autosomal microsatellite markers, and explored the genetic diversity and the measure of population structuring of Eurasian Woodcock (Scolopax rusticola) in Hungary in spring. We tested whether any sub-populations can be identified in our sample, and we also examined whether the individuals that occurred closer to each other in space and in time were also more similar in genetic terms. We analysed samples from 240 free-ranging birds collected during legal hunting from mid-February to the end of April in 2015 from different parts of the country. The Bayesian clustering method and Markov Chain Monte Carlo simulation were used to infer the most probable number of genetic clusters without a priori definition of populations. The second approach was a discriminant analysis of principal components in order to identify clusters of individuals without population genetic models. Additionally, we fitted a general linear model with the genetic distance between samples as the dependent variable, the temporal (days) and the geographical (meters) distances and the interactions of these factors as independent variables. We found high genetic diversity and low level of population structuring in our samples. Moreover, our results did not support the assumptions that Woodcocks occurring in different places or at different times in Hungary would also belong to different breeding populations.

1. Introduction

The Eurasian Woodcock (Scolopax rusticola) is a widespread wader, occurring over most of the Palearctic (Ferrand & Gossmann 2009a). It is a migratory species across most of its range, the majority of the birds move between different breeding and wintering areas with some individuals travelling thousands of kilometres (Guzmán et al. 2011, Arizaga et al. 2015). In the western Palearctic, northern and eastern Europe are considered to be the main breeding areas, while western and southern Europe are the main wintering areas (Ferrand & Gossmann 2009a). There are no definite borders between these sites, moreover Woodcock also breed in western Europe (Ferrand et al. 2008,
Braña et al. 2013, Heward et al. 2015). The density of birds at wintering areas can be much higher than in the much larger breeding areas, and Woodcock from different origins mix at those places (Guzmán et al. 2011, Iljinsky et al. 2013, Arizaga et al. 2015).

Hungary, or more generally the Carpathian Basin, can be considered as a transition zone between the main breeding and wintering areas of the species. Woodcocks occur in the country during the whole year, but their abundance is higher during spring and autumn, which is most likely related to migration (Szemethy et al. 2014). The local breeding population is generally thought to be small (several hundred individuals (BirdLife International 2015)), but only poor quality data are currently available.

In the case of migratory species, understanding of connectivity patterns between breeding-, wintering- and passage areas is fundamental for effective conservation. However, to date, there is still little information on the origins of the individuals that can be observed in Hungary. On the basis of ringing data (Schally 2015), we assumed that Woodcock belonging to different breeding populations may migrate across this region. Ringing data of probable nesting sites are relatively few, yet they cover a wide geographical range, from the circum-Baltic to the east of Moscow region. Furthermore, as former studies based on ring-recovery data (Guzmán et al. 2011), stable isotope analyses (Hobson et al. 2013) and telemetry data (Arizaga et al. 2015) provided generally consistent yet somewhat different result regarding the origin areas of a given population, it is likely that some connections to breeding areas may still have remained undetected.

Therefore, our aim was to expand the knowledge available with basic information about the potential number of different breeding areas that can be in connection with the country, using molecular genetic tools. We also investigated probable spatial and temporal patterns of genetic relationships among the birds. We assumed that the individuals moving to similar destinations might migrate at similar times and via similar routes, and therefore there might be a spatial or temporal pattern in the occurrences of such groups.

In this study, we examined the variation of eight microsatellite markers, and explored the genetic diversity and the measure of population structuring of Eurasian Woodcock in Hungary in the spring migration period. We tested whether any subpopulations can be identified in the sample, and we also examined whether the individuals that occurred closer to each other in space and in time were also more similar in genetic terms.

### 2. Material and methods

#### 2.1. Sample collection and DNA extraction

We analysed tissue samples from 240 Eurasian Woodcocks. Muscle tissues were obtained from free-ranging birds collected during legal hunting by hunters collaborating with “Woodcock Monitoring Programme” (Szemethy et al. 2014) in Hungary between 15th February and 30th April 2015. In order to maximize representativity, the following aspects were taken into account. The country was divided into four parts according to its geographical center (NW, NE, SW, SE), and the period of sample collection was divided into 3 pe-

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First-year</td>
<td>Adult</td>
<td>First-year</td>
</tr>
<tr>
<td>Period 1 Male</td>
<td>8</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Female</td>
<td>3</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Period 2 Male</td>
<td>12</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Female</td>
<td>5</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>Period 3 Male</td>
<td>8</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Female</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 1. The number of samples grouped by sampling area, -period, sex and age of the birds.
periods of similar length (Period 1: 15$_{th}$ February–7$_{th}$ March, period 2: 8$_{th}$ March–5$_{th}$ April, period 3: 6$_{th}$ April–30$_{th}$ April). These periods were also adjusted according to the formerly described characteristics of Woodcock detections (increasing-, peaking- and decreasing phases) in the country (Schally et al. 2012).

We have analysed a similar number of samples from each area and each time period. Moreover, these groups involved as equal rate as possible of birds according to sex and age (Table 1). Sex of the birds was determined by dissection of the gonads, while the age was determined by the study of the moult stages according to Ferrand & Gossmann (2009b). Five birds were not able to be aged properly. These groups provided only the representation of the main characteristics of the population, but they were not used as factors for further analyses.

Muscle samples were stored in ethanol at –20 °C. Total genomic DNA was extracted using the commercial Genomic DNA Mini Kit (Geneaid Biotech Ltd., Taiwan) according to the manufacturers instructions. After extraction, genomic DNA was stocked at –20 °C until processing.

2.2. Microsatellite genotyping

Samples were genotyped by PCR amplification at 8 microsatellite loci (Sru1-24b, Sru1-24c, Sru54b, Sru74a, Sru79d, Sru87b, Sru113a, Sru128b) that had previously been isolated and tested in the species (Cardia et al. 2007). Samples were amplified in two multiplex reactions using QIAGEN Multiplex PCR Kit (QIAGEN GmbH, Germany) in a final volume of 25 µl. Microsatellite genotyping was performed on an ABI Prism 3100 Genetic Analyser with LIZ500 Size Standard (Applied Biosystems, USA). Negative controls were included for amplification procedures. Allele sizes were scored with PeakScanner ver. 1.0 (Applied Biosystems, USA).

Microchecker ver. 2.2.3 (van Oosterhout et al. 2004) was employed to search for null alleles, scoring errors and large allele dropout. In order to avoid resampling of individuals, the Identity Analysis of the CERVUS ver. 3.0.6 (Kalinowski et al. 2007) was carried out. The number of alleles per locus (Na), the expected ($H_e$) and observed heterozygosity values ($H_o$), deviations from Hardy–Weinberg equilibrium after Bonferroni correction (HWE), and measures of genetic diversity for each locus and averaged across loci were calculated with CERVUS software (Kalinowski et al. 2007) and GenAlEx ver. 6.501 (Peakall & Smouse 2012).

2.3. Assessing genetic structure

Several different approaches were used to assess population differentiation in the study area. The Bayesian clustering method and Markov Chain Monte Carlo (MCMC) simulation implemented in STRUCTURE ver. 2.3.4 (Pritchard et al. 2000) were used to infer the most probable number of genetic clusters without a priori definition of populations. The analyses were run using an admixture model and correlated allele frequencies with a burn-in period of 250,000 replicates and a sampling period of 750,000 replicates for number of clusters (K) from 1 to 10 with ten independent runs for each K. To determine the number of genetic clusters, we used the method of Evanno et al. (2005) based on the second order rate of change in log Pr (X|K) as implemented by the program Structure Harvester ver. 0.6.94 (Earl & VonHoldt 2012).

The second approach was a discriminant analysis of principal components (DAPC), a multivariate method implemented in the adegenet package (Jombart 2008) with R ver. 3.3.1 (R Development Core Team) that identifies clusters of individuals without using any population genetic model (Jombart et al. 2010). We used the “find.clusters” function for the identification of the optimal number of clusters (K) with the “choose.n.clust” option and the Bayesian Information Criterion (BIC). After that, DAPC was employed to assign individuals into populations, retaining all the principal components, as suggested in the manual. Additionally, to detect possible spatio-temporal patterns of population structuring, we performed a general linear model with the genetic distance between samples as the dependent variable, the temporal (days) and the geographical (meters) distances and the interactions of these factors as independent variables. Genetic distance values were calculated with GenAlEx, and the model was fitted.
Table 2. Number of alleles (Na), number of effective alleles (Ne), observed and expected heterozygosity values (Ho and He, respectively), deviation from Hardy–Weinberg proportions (**p < 0.005) after Bonferroni correction (HWE), polymorphic information content (PIC), and Shannon’s information index (I) in the sampled Eurasian Woodcock.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Na</th>
<th>Ne</th>
<th>Ho</th>
<th>He</th>
<th>HWE</th>
<th>PIC</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sru1-24b</td>
<td>15</td>
<td>6.85</td>
<td>0.788</td>
<td>0.856</td>
<td>NS</td>
<td>0.838</td>
<td>2.142</td>
</tr>
<tr>
<td>Sru1-24c</td>
<td>4</td>
<td>2.19</td>
<td>0.548</td>
<td>0.545</td>
<td>NS</td>
<td>0.471</td>
<td>0.921</td>
</tr>
<tr>
<td>Sru54b</td>
<td>13</td>
<td>6.78</td>
<td>0.616</td>
<td>0.854</td>
<td>***</td>
<td>0.836</td>
<td>2.080</td>
</tr>
<tr>
<td>Sru74a</td>
<td>9</td>
<td>4.06</td>
<td>0.639</td>
<td>0.753</td>
<td>NS</td>
<td>0.722</td>
<td>1.663</td>
</tr>
<tr>
<td>Sru79d</td>
<td>10</td>
<td>3.69</td>
<td>0.718</td>
<td>0.730</td>
<td>NS</td>
<td>0.699</td>
<td>1.599</td>
</tr>
<tr>
<td>Sru87b</td>
<td>4</td>
<td>2.22</td>
<td>0.372</td>
<td>0.551</td>
<td>***</td>
<td>0.485</td>
<td>0.937</td>
</tr>
<tr>
<td>Sru113a</td>
<td>5</td>
<td>1.49</td>
<td>0.339</td>
<td>0.330</td>
<td>NS</td>
<td>0.309</td>
<td>0.674</td>
</tr>
<tr>
<td>Sru128b</td>
<td>9</td>
<td>2.56</td>
<td>0.657</td>
<td>0.611</td>
<td>NS</td>
<td>0.552</td>
<td>1.142</td>
</tr>
<tr>
<td>Mean</td>
<td>8.625</td>
<td>3.73</td>
<td>0.585</td>
<td>0.654</td>
<td>–</td>
<td>0.614</td>
<td>1.395</td>
</tr>
</tbody>
</table>

Fig. 1. A) The mean log-likelihood values for each value of the number of clusters. The most likely number of clusters is where the Ln probability is maximized. B) The probability of the models according to cluster sizes. The most likely number of clusters is where Delta K is the highest. C) The cluster assignment of the individuals in STRUCTURE. The bars represent the individuals, the colours indicate their assignment to different clusters in the case of 5 clusters (K = 5).
using the function “lm” with R ver. 3.3.1 (R Development Core Team).

3. Results

3.1. Genetic diversity

Genetic variability was relatively high, the mean number of alleles reached 8.625, and ranged between 4 and 15 per locus (Table 2). Mean observed and expected heterozygosities were 0.585 and 0.654, respectively. Significant deviations of Hardy–Weinberg proportions were found in two out of the eight loci; deviations were caused by heterozygote deficit, probably due to high frequencies of null alleles. Polymorphic Information Content (PIC) ranged between 0.309 and 0.838 with a mean value of 0.614, Shannon’s Information Index ranged between 0.674 and 2.142 with a mean value of 1.395, both showing a relatively high genetic diversity.

3.2. Genetic structure

Genetic structure was detected both by STRUCTURE and DAPC analyses. STRUCTURE assigned the highest average likelihood scores to the cases of a unique genetic unit ($K = 1$) in the whole dataset, whereas the second order rate of change in log Pr indicate the presence of more clusters, but even forcing this higher number of clusters ($K = 5$), our results did not show any clear groups, only mixed genotypes (Fig. 1).

The DAPC with the “find.clusters” function also revealed the presence of genetic subunits. The lowest BIC values were obtained for a model comprising eight clusters, but they were all very similar for 6–8 clusters (Fig. 2).

The general linear model showed significant, but very weak relationship between the genetic distances and the spatio-temporal distances (multiple $R^2 = 0.002$; $F_{3,28} = 21.48; p < 0.001$).

4. Discussion

So far, few studies have investigated the genetic aspects of the Eurasian Woodcock. Burlando et al. (1996) used random amplification of polymorphic DNA (RAPD) markers for the study of genetic diversity, and have shown that genetic variation was higher within than between populations. Memoli & Paffetti (2007) also used RAPD markers with mitochondrial sequences to assess genetic diversity of Italian birds. These, and subsequent mitochondrial sequencing studies (Trucchi et al. 2011), did not show any phylogenetic patterns in the species.
Relatively high genetic diversity was found in our sample of the population using microsatellites. According to the PIC, which is an efficient indicator of loci adequacy in population genetics studies, the loci optimized in this study can be considered moderate to highly informative (Dawny et al. 2009, Souza et al. 2012). The potential of our set of markers for identifying individuals within the population was quite high, i.e., combined probability of identity across loci was $1 \times 10^{-6}$. This relatively high probability, combined with the high information content of the markers indicates that this panel of microsatellites is useful not only in the estimation of the overall level of genetic variability in Eurasian Woodcock populations, but also in the diagnosis of relationships between individuals (Dawny et al. 2009, Szabolcsi et al. 2014).

Observed heterozygosity values in the Eurasian Woodcock (0.339–0.788) per locus were in the range noted in the species previously (Cardia et al. 2007). Overall heterozygosity ($H_o = 0.585$) was around the reported average value for 76 species of birds (Eo et al. 2011). The mean number of alleles per locus ($N_a = 8.625$) was higher than the average for the former mentioned species examined by Eo et al. (2011), even when considering only the terrestrial bird species.

According to the weak differentiation found in our samples, grouping all of them in one cluster seems more appropriate, suggesting an extensive admixture between populations of different nesting sites. Although the DAPC indicated some structuring, the established clusters greatly overlapped and could not be separated clearly. This was also in accordance with the membership probability results of STRUCTURE. Various demographic and historical factors may contribute to the lack of population structure. High level of dispersal would be enough to prevent genetic structuring, as it tends to oppose the effect of genetic drift and homogenize populations (Slatkin 1989). There is no exact information about philopatry and the extent of natal dispersal in the species, but ringing data indicate that some individuals may breed away from their natal site (Hoodless & Coulson 1994, Schally 2015).

In particular at wintering grounds, there is high potential for mixing between birds of different regions. Some winter recoveries in the Mediterranean indicate a substantial overlap of the wintering areas among birds from different breeding regions (Ferrand et al. 2008, Guzmán et al. 2011, Arizaga et al. 2015). Ramenofsky and Wingfield (2006) suggested that some stages of migration and breeding may overlap in long distance migrant birds. If mating of Woodcock is not limited to the breeding sites, but occurs also during migration, this could affect population structuring like dispersal. Male and female Woodcock originating from different natal sites, regardless of their philopatry, could mate on migration and parent offspring together. Previous reports have shown that during the migration period both male and female birds were in their early stage of sexual activity (Elbling et al. 2008), but the details of this phenomenon should be studied further.

Even though some genetic structuring could be expected given the large breeding area and proposed nest-site fidelity in the Eurasian Woodcock, the weak differentiation found in this study is similar to results from previous reports where little or no genetic differentiation between populations has been found for the species (Burlando et al. 1996, Trucchi et al. 2011) and other migratory birds, like the Tufted Duck (Aythya fuligula) (Liu et al. 2012, 2013), the Mallard (Anas platyrhynchos) (Kraus et al. 2013), the Greylag Goose (Anser anser) (Pellegrino et al. 2015), the Barnacle Goose (Branta leucopsis) (Jonker et al. 2013), the Little Auk (Alle alle) (Wojcuzanis-Jakubas et al. 2014), the Crested Auklet (Aethia cristatella) (Pshenichnikova et al. 2015), the Grey-headed Albatross (Thalassarche chrysostoma) (Burg & Croxall 2001), and the Baillon’s Crane (Zapornia pusilla) (Seifert et al. 2016).

The high mobility of long-distance migrants can have strong effect on their population structure, as it may allow dispersal and gene flow at macrogeographic scales. In the most extreme case, an entire species could then be composed of a single panmictic population (Liu et al. 2012). Population subdivision in spite of high mobility may exist if individuals are philopatric and consistently return to the same breeding grounds. In fact, evidence of different subpopulations characterized by some genetic divergence is available from several long-distance migrants, like the Dunlin (Calidris alpina) (Wennerberg, 2001) or Swainson’s Thrush (Catharus ustulatus) (Ruegg & Smith, 2002).

The low level of structuring can also be ex-
plained by the continuous nature of European breeding populations and the lack of barriers. If any population structuring does occur, it is not more likely to represent gradual differentiation over very large distances, and perhaps, this cannot be observed in a small snapshot of birds, like the ones that pass through the Hungarian flyway.

Our results did not support the assumptions that Woodcocks occurring in different places at different times in Hungary would belong to different breeding populations. The reason for a lack of spatial or temporal patterns in genetic differences is clear if there are no subpopulations to be distinguished. However, if there is still some structuring remaining undetected, this result needs further explanation. Birds with different breeding areas may have different journey lengths or flight paths in order to optimize their arrival, however high level of individual variation can still cause large temporal or spatial overlaps among them during their migration. Additionally, individuals can also alter their migration routes according to locally or regionally variable environmental conditions.

According to the European breeding population estimates (BirdLife International 2015), and also to the European hunting bag estimates (Ferrand & Gossmann 2001) it is possible that a population in the scale of millions of individuals may traverse through the central European region during spring. Population genetic studies of such big populations would also require very large sample sizes in order to be able to identify differences clearly. For further studies about migration connectivity, it would be essential to compare our samples to populations of a broad range of breeding areas.

Acknowledgments. We are very grateful to every hunter for participation in sample collection. We are also thankful to the Hungarian Hunters’ National Association and the Ministry of Agriculture for their help in coordination of the monitoring programme. The research was supported by the Research Centre of Excellence (1476-4/2016/FEKUT).

References


Burg, T.M. & Croxall, J.P. 2001: Global relationships amongst black-browed and grey-headed albatrosses:


